

REMARKS

In the Office Action, the Examiner rejected claims 10, 11, 20, and 21 under 35 U.S.C. § 101 as directed to non-statutory subject matter; rejected claims 1-3, 8-14, 20, and 21 under 35 U.S.C. § 102(b) as anticipated by NAJORK et al. (U.S. Patent No. 6,321,265); and rejected claims 4-7, 15-19, and 22-24 under 35 U.S.C. § 103(a) as unpatentable over NAJORK et al. Applicants respectfully traverse these rejections.

By the present amendment, Applicants amend claims 1, 7, 10, 12, and 18-20 to improve form and add new claim 25. Claims 1-25 are pending.

Claims 10, 11, 20, and 21 were rejected under 35 U.S.C. § 101 as allegedly directed to non-statutory subject matter. In particular, the Examiner alleged that "[d]ependent claims 11 and 21 include a 'carrier wave' which is non-statutory because it does not fit into any of the three product statutory classes because it is non-physical" (Office Action, pg. 2). Applicants respectfully traverse.

The Examiner has provided no basis for rejecting claims 10, 11, 20, and 21 under 35 U.S.C. § 101. The mere fact that claims 11 and 21 recite a carrier wave, which is a type of computer-readable medium, in no way means that these claims are per se non-statutory. The Examiner's general allegation is insufficient for establishing a *prima facie* basis for denying patentability. If this rejection is maintained, Applicants request that the Examiner specifically point out the basis for the rejection so that Applicants can address the Examiner's concerns.

Applicants also note that the Examiner alleged that claims 10 and 20, from which claims 11 and 21 depend, are directed to non-statutory subject matter "because they are

broader than claims 11 and 21" (Office Action, pg. 3). The Examiner has not provided a prima facie basis for denying patentability. The mere fact that claims 10 and 20 are broader than claims 11 and 21 is in no way a per se indication that these claims are non-statutory. If the Examiner maintains the rejection of claims 10, 11, 20, and 21 under 35 U.S.C. § 101, Applicants request that the Examiner specifically point out why the Examiner alleges that the features recited in claims 10 and 20 are directed to non-statutory subject matter.

For at least the foregoing reasons, Applicants request that the rejection of claims 10, 11, 20, and 21 under 35 U.S.C. § 101 be reconsidered and withdrawn.

Claims 1-3, 8-14, 20, and 21 were rejected under 35 U.S.C. § 102(b) as allegedly anticipated by NAJORK et al. Applicants respectfully traverse this rejection.

At the outset, Applicants note that the ground of rejection is improper. Applicants' application was filed on August 14, 2000, and claims priority to U.S. Provisional Application No. 60/195,581, which was filed on April 6, 2000. NAJORK et al. was issued on November 20, 2001, after the August 14, 2000, filing date of Applicants' application. Applicants assume that the Examiner intended to reject claims 1-3, 8-14, 20, and 21 under 35 U.S.C. § 102(e) and not 102(b).

A proper rejection under 35 U.S.C. § 102 requires that a single reference teach every aspect of the claimed invention either explicitly or impliedly. Any feature not directly taught must be inherently present. See M.P.E.P. § 2131. NAJORK et al. does not disclose or suggest the combination of features recited in claims 1-3, 8-14, 20, and 21.

Amended independent claim 1, for example, is directed to a computer implemented method of crawling hyperlinked documents. The method includes sending a request for additional links to hyperlinked documents to a link manager, receiving a plurality of links to hyperlinked documents to be crawled, where the plurality of links is selected by the link manager based on priority, grouping the plurality of links to hyperlinked documents by host, selecting a host to crawl next according to a stall time of the host, and crawling a hyperlinked document from the selected host. NAJORK et al. does not disclose or suggest this combination of features.

For example, NAJORK et al. does not disclose or suggest receiving a plurality of links to hyperlinked documents to be crawled, where the plurality of links is selected by the link manager based on priority. The Examiner relied on col. 1, lines 31-47, and col. 3, lines 3-52, of NAJORK et al. for allegedly disclosing receiving a plurality of links to hyperlinked documents to be crawled (Office Action, pg. 4). Applicants submit that neither these sections nor elsewhere in NAJORK et al. discloses or suggests that the received plurality of links to hyperlinked documents are links that are selected by a link manager based on priority.

At col. 1, lines 31-47, NAJORK et al. discloses:

A web crawler is a program that automatically finds and downloads documents from host computers in networks such as the world wide web. When a web crawler is given a set of starting URL's, the web crawler downloads the corresponding documents, then the web crawler extracts any URL's contained in those downloaded documents and downloads more documents using the newly discovered URL's. This process repeats indefinitely or until a predetermined stop condition occurs. As of 1999 there were approximately 500 million web pages on the world wide web, and the number is continuously growing; thus, web crawlers need efficient data structures to keep track of downloaded documents and any discovered

addresses of documents to be downloaded. One common data structure to keep track of addresses of documents to be downloaded is a first-in-first-out (FIFO) queue. Using FIFO queues, URL's are enqueued as they are discovered, and dequeued in the order enqueued when the crawler needs a new URL to download.

This section of NAJORK et al. discloses that a web crawler is given a set of starting URLs. This section of NAJORK et al. does not disclose or suggest that the set of starting URLs are selected by a link manager based on priority. In fact, this section of NAJORK et al. in no way discloses or suggests that some URLs are higher priority than other URLs.

At col. 3, lines 3-52, NAJORK et al. discloses that a web crawler enqueues URLs in a set of queues, with all URLs sharing a respective common host address being stored in a respective common one of the queues. This section of NAJORK et al. does not disclose or suggest that the URLs are selected by a link manager based on priority. In fact, this section of NAJORK et al. in no way discloses or suggests that some URLs are higher priority than other URLs.

Since NAJORK et al. does not disclose receiving a plurality of links to hyperlinked documents to be crawled, where the plurality of links are selected by the link manager based on priority, Applicants respectfully submit that the rejection of claim 1 under 35 U.S.C. § 102(e) as anticipated by NAJORK et al. should be reconsidered and withdrawn.

For at least the foregoing reasons, Applicants submit that claim 1 is not anticipated by NAJORK et al.

Claims 2, 3, 8, and 9 depend from claim 1. Therefore, these claims are not anticipated by NAJORK et al. for at least the reasons given above with respect to claim 1.

Amended independent claims 10, 12, and 20 recite features similar to the features recited above with respect to claim 1. Therefore, Applicants submit that these claims are not anticipated by NAJORK et al. for reasons similar to the reasons given above with respect to claim 1.

Claim 11 depends from claim 10. Therefore, this claim is not anticipated by NAJORK et al. for at least the reasons given above with respect to claim 10.

Claims 13 and 14 depend from claim 12. Therefore, these claims are not anticipated by NAJORK et al. for at least the reasons given above with respect to claim 12.

Claim 21 depends from claim 20. Therefore, this claim is not anticipated by NAJORK et al. for at least the reasons given above with respect to claim 20.

Claims 4-7, 15-19, and 22-24 were rejected under 35 U.S.C. § 103(a) as unpatentable over NAJORK et al. Applicants respectfully traverse this rejection.

Claims 4-7 depend from claim 1. Accordingly, Applicants submit that claims 4-7 are patentable over NAJORK et al. for at least the reasons given above with respect to claim 1. Moreover, these claims recite additional features that are neither disclosed nor suggested by NAJORK et al.

For example, claim 4 recites grouping the hosts according to the number of hyperlinked documents to be crawled at each host. The Examiner alleged that NAJORK

et al. discloses grouping hosts and pointed to Fig. 7 and col. 2, lines 24-36, of NAJORK et al. for support (Office Action, pg. 7). Applicants disagree.

Fig. 7 of NAJORK et al. illustrates a host-to-queue assignment table 132 that is used according to the second embodiment of NAJORK et al.'s system. As clearly illustrated in Fig. 7, each host is assigned to a different queue. This figure of NAJORK et al. in no way supports the Examiner's allegation. That is, this figure of NAJORK et al. does not disclose or suggest the grouping of hosts, but rather the assignment of hosts to different queues.

At col. 2, lines 24-36, NAJORK et al. discloses:

The Scooter web crawler used by AltaVista uses a different approach. Scooter keeps a first list of URL's of web pages to be downloaded, and a second list of host computers from which downloads are in progress. Newly discovered URL's are added to the end of the first list. To locate a new URL to download, Scooter compares items in the first list with the second list until it finds a URL whose host computer is not in the second list. Scooter then removes that URL from the first list, updates the second list, and downloads the corresponding document. One of the disadvantages of this approach is the time wasted scanning through the list of URL's each time a thread in the crawler is ready to perform a download.

This section of NAJORK et al. discloses that the Scooter web crawler keeps a list of host computers from which downloads are in progress. In such a scenario, there would be no need to group host computers since all host computers from which downloads are in progress are placed on the list. Contrary to the Examiner's allegation, this section of NAJORK et al. does not disclose or suggest grouping hosts.

Even assuming, for the sake of argument, that NAJORK et al. could reasonably be construed to disclose grouping hosts, NAJORK et al. does not disclose or suggest that the grouping of hosts occurs in accordance to the number of hyperlinked documents to be

crawled at each host. The Examiner admits that NAJORK et al. does not disclose this feature and alleged "[i]t would have been obvious and desirable to have grouped the hosts according to the number of hyperlinked documents to be crawled so that the largest groups could have been processed first" (Office Action, pg. 7). Applicants submit that the rejection of claim 4 is improper.

A proper rejection under 35 U.S.C. § 103(a) requires that the Examiner set forth in the Office Action (1) the relevant teachings of the prior art reference(s) relied upon, preferably with reference to the relevant column or page number(s) and line number(s) where appropriate, (2) the difference or differences in the claim over the applied reference(s), (3) the proposed modification of the applied reference(s) necessary to arrive at the claimed subject matter, and (4) an explanation why one of ordinary skill in the art at the time the invention was made would have been motivated to make the proposed modification. See M.P.E.P. § 706.02(j).

The Examiner does not logically explain why modifying NAJORK et al. to include grouping hosts according to the number of hyperlinked documents to be crawled at each host, as required by claim 4, would allow the largest groups to be processed first. As such, the Examiner has not logically explained why one of ordinary skill in the art at the time the invention was made would have been motivated to make the proposed modification. Accordingly, the rejection of claim 4 under 35 U.S.C. § 103(a) is improper.

Moreover, the Examiner's allegation is merely conclusory and insufficient for establishing a *prima facie* case of obviousness.

For at least these additional reasons, Applicants submit that claim 4 is patentable over NAJORK et al.

Claim 5 recites examining the groups in descending order of the number of hyperlinked documents to be crawled at each host until a host is found with a stall time that is earlier than the current time. The Examiner alleged that NAJORK et al. discloses this feature and relied on Figs. 5-7, col. 1, line 31, to col. 2, line 2, and col. 2, lines 37-62, of NAJORK et al. for support (Office Action, pg. 7).

At the outset, Applicants submit that since the Examiner admitted that NAJORK et al. does not disclose grouping hosts according to the number of hyperlinked documents to be crawled at each host (Office Action, pg. 7), NAJORK et al. cannot, contrary to the Examiner's allegation, disclose examining the groups in descending order of the number of hyperlinked documents to be crawled at each host until a host is found with a stall time that is earlier than the current time. If this rejection is maintained, Applicants request that the Examiner explain how NAJORK et al. can disclose a feature that the Examiner admitted that NAJORK et al. does not disclose.

Fig. 5 of NAJORK et al. illustrates a process for dequeuing a URL to be crawled. This figure does not disclose or suggest examining the groups in descending order of the number of hyperlinked documents to be crawled at each host until a host is found with a stall time that is earlier than the current time, as required by claim 5.

Fig. 6 of NAJORK et al. illustrates the demultiplexing of URLs from a main first-in, first-out (FIFO) queue 242 to individual FIFO queues 246-1 to 246-n. This figure does not disclose or suggest examining the groups in descending order of the number of

hyperlinked documents to be crawled at each host until a host is found with a stall time that is earlier than the current time, as required by claim 5.

Fig. 7 of NAJORK et al. illustrates a host-to-queue assignment table 132. This figure does not disclose or suggest examining the groups in descending order of the number of hyperlinked documents to be crawled at each host until a host is found with a stall time that is earlier than the current time, as required by claim 5.

At col. 1, line 31, to col. 2, line 2, NAJORK et al. discloses that a prior art web crawler downloads documents, extracts any URLs from the downloaded documents, and downloads more documents based on the extracted URLs. This section of NAJORK et al. in no way discloses or suggests examining the groups in descending order of the number of hyperlinked documents to be crawled at each host until a host is found with a stall time that is earlier than the current time, as required by claim 5.

At col. 2, lines 37-62, NAJORK et al. discloses that the prior art Scooter web crawler implements a politeness policy so that host computers are not overloaded by requests from the Scooter web crawler. This section of NAJORK et al. in no way discloses or suggests examining the groups in descending order of the number of hyperlinked documents to be crawled at each host until a host is found with a stall time that is earlier than the current time, as required by claim 5.

For at least these additional reasons, Applicants submit that claim 5 is patentable over NAJORK et al.

Claims 15-19 depend from claim 12. Accordingly, Applicants submit that claims 15-19 are patentable over NAJORK et al. for at least the reasons given above with respect

to claim 12. Moreover, these claims recite additional features that are neither disclosed nor suggested by NAJORK et al.

For example, claims 15-19 recite features similar to features recited above with respect to claims 4-7. Therefore, Applicants submit that claims 15-19 are further patentable over NAJORK et al. for reasons similar to the reasons given above with respect to claims 4-7.

Independent claim 22 is directed to a computer implemented method of crawling hyperlinked documents. The method includes storing a plurality of links to hyperlinked documents to be crawled; determining that more links to hyperlinked documents are desired; sending requests to multiple link managers for more links to hyperlinked documents; receiving additional links to hyperlinked documents from the link managers; selecting a host to crawl next according to a stall time of the host; and crawling a hyperlinked document from the selected host. NAJORK et al. does not disclose or suggest this combination of features.

For example, NAJORK et al. does not disclose or suggest sending requests to multiple link managers for more links to hyperlinked documents. The Examiner admitted that NAJORK et al. does not disclose this feature and alleged that "[i]t would have been obvious ... to have modified the queues and indexing system of Najork to have operated as a link manager from which the web crawler could have requested more links to hyperlinked documents" (Office Action, pp. 9-10). Applicants submit that the Examiner's allegation does not address the above feature of claim 22.

Claim 22 recites sending requests to multiple link managers for more links to hyperlinked documents. Therefore, even if, as alleged by the Examiner, NAJORK et al.'s queues and indexing system could reasonably operate as a link manager, the Examiner did not logically explain how or why one skilled in the art would seek to modify NAJORK et al.'s system to include multiple link managers to which requests for more links to hyperlinked documents can be made. Accordingly, the Examiner did not establish a *prima facie* case of obviousness with respect claim 22.

NAJORK et al. discloses that indexing system 116 includes an index of words used on the world wide web and addresses of the web pages that use each word (col. 4, lines 63-65). NAJORK et al. also discloses that web crawler 102 can access indexing system 116 in the process of downloading web pages from the world wide web (col. 4, lines 67, to col. 5, line 3). NAJORK et al. does not disclose or suggest, however, that indexing system 116 receives requests for more links to hyperlinked documents or sends additional links to hyperlinked documents. Therefore, the Examiner's allegation that indexing system 116 can somehow be construed to be a link manager is unsupported by the NAJORK et al. disclosure.

For at least the foregoing reasons, Applicants submit that claim 22 is patentable over NAJORK et al.

Independent claim 23 recites features similar to the features recited above with respect to claim 22. Therefore, Applicants submit that claim 23 is patentable over NAJORK et al. for reasons similar to the reasons given above with respect to claim 22.

Claim 24 depends from claim 23. Therefore, Applicants submit that this claim is patentable over NAJORK et al. for at least the reasons given above with respect to claim 23.

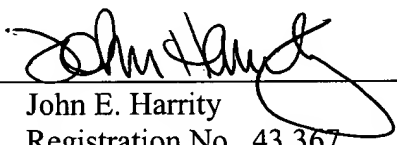
New independent claim 25 is directed to a method for crawling hyperlinked documents. The method includes prioritizing a plurality of links to hyperlinked documents to be crawled and crawling a hyperlinked document using one of the prioritized plurality of links. NAJORK et al. does not disclose or suggest this combination of features.

In view of the foregoing amendment and remarks, Applicants respectfully request the Examiner's reconsideration of this application, and the timely allowance of the pending claims.

To the extent necessary, a petition for an extension of time under 37 C.F.R. § 1.136 is hereby made. Please charge any shortage in fees due in connection with the filing of this paper, including extension of time fees, to Deposit Account No. 50-1070 and please credit any excess fees to such deposit account.

Respectfully submitted,

HARRITY & SNYDER, L.L.P.

By: 
John E. Harrity
Registration No. 43,367

Date: May 12, 2004

11240 Waples Mill Road
Suite 300
Fairfax, Virginia 22030
(571) 432-0800

Customer Number: 26615